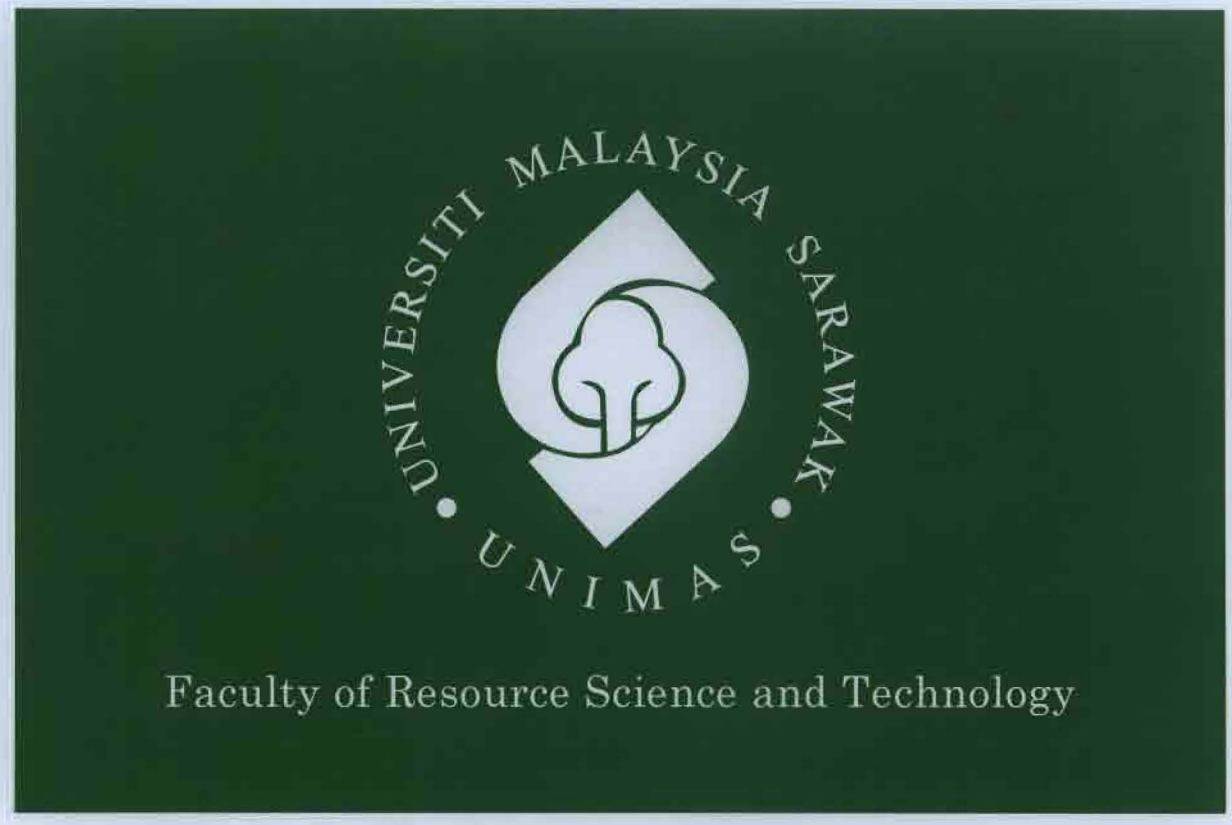

Prediction of Logical Interactions between Ribosomal Protein L22 with the Factors of Epstein Barr Virus Genes using *In Silico* Strategy

Athma Hafeeza Binti Marzuki
(34779)

Bachelor of Science with Honours
(Resource Biotechnology)
2015

Prediction of Logical Interaction between Ribosomal Protein L22 with the Factors of EBV Genes using *In Silico* Strategy

Athma Hafeeza binti Marzuki (34779)

A Thesis Submitted in Partial Fulfillment of the Requirement of the Degree of Bachelor Science with Honours (Resource Biotechnology)

Supervisor: Assoc. Prof. Dr. Edmund Sim Ui Hang

Resource Biotechnology
Department of Molecular Biology

Faculty of Resource Science and Technology
Universiti Malaysia Sarawak

Acknowledgement

First of all, I would like thank the God who granted me the patience and strength to complete this project to a success. I would also like to express my highest gratitude to my project supervisor, Associate Professor Dr. Edmund Sim Ui Hang who has been supervising and guiding me throughout the duration of my project as well as for giving me advice tips to conduct this final year project to a successful end. I hereby thank him for always being a support throughout the completion of my final year project.

Furthermore, I would like to pay my highest appreciation to the post graduate students of the Immunological Human Molecular Genetics Laboratory, UNIMAS. Particularly, Ms. Stella Chan Li Li, Ms. Kherlee Ng, and Ms. Shruti Talwar who have been consistently guiding me by sharing their knowledge and generously helped and discussed my project throughout the process of completion of my project. In addition, I would like appreciate my lab mates, namely Najian, Nur Suriati, Yew Keh Li, Cassandra Chee Sheau Mei, Lisha, Nazatul Syahira and Jaiyogesh for their kindness, support and unity to make my project a success.

Last but not least, i want to thank my mother, Mrs. Rukiah and my father, Mr. Marzuki for their endless support and encouragement throughout the duration of my final year project. Their moral support and their trust on me have always been the bridge for the success of my project. In addition, I would like to thank all my friends and coursemates for supporting and encouraging me throughout the completion of my project.

Declaration

I hereby declare that this thesis entitled "Predicting Logical Interaction between Ribosomal Protein L22 and the Factors from EBV Genes using *In Silico* Strategy" is my own work and all sources ave been quoted and referred to and have been acknowledged by means of complete references. It has been submitted to Universiti Malaysia Sarawak and shall not be submitted to other universities or institute of higher learning.

.....

(ATHMA HAFEEZA BINTI MARZUKI)

Date:

Table of contents

List of Abbreviations

EBV	Epstein Barr virus
NPC	Nasopharyngeal Carcinoma
RPL 22	Ribosomal protein L22
RNA	Ribonucleic Acid
EBNA	Epstein Barr nuclear antigen
EBER	Epstein Barr encoded RNAs
LMP	Latent membrane protein
PDB	Protein Data Bank
Arg	Arginine
Lys	Lysine
Asn	Asparagine
Ile	Isoleucine
Leu	Leucine
Thr	Threonine
Trp	Tryptophan
His	Histidine
Ala	Alanine
Val	Valine
Ser	Serine
Asp	Aspartic Acid
Glu	Glutamic Acid
Gln	Glutamine
Gly	Glycine
Tyr	Tyrosine

List of Tables

List of Figures

Prediction of Logical Interactions between Ribosomal Protein L22 with the factors of EBV genes using *In silico* Strategy

Athma Hafeeza binti Marzuki (34779)

Resource Biotechnology
Faculty of Resource Science and Technology
Universiti Malaysia Sarawak

ABSTRACT

Epstein Barr virus (EBV) was first discovered in Burkitt-Lymphoma cells. It was later found in nasopharyngeal cancer (NPC) cells, which attracts the attention of many researchers. EBV genes produces some factors when the genes are expressed. The factors produced include basic tegument proteins, glycoproteins, capsid proteins, nuclear antigens as well as RNAs. Previous studies show that EBER 1, one of these factors somehow interact with ribosomal protein L22, which was found in the 60S large ribosomal subunit of the ribosomes. Its function remains unknown, but somehow its association with one of the factors from EBV genes was believed to be the cause of cancer. Thus, this study was conducted to perform the prediction of interaction between RPL22 with the factors of EBV genes using bioinformatics tools, and at which region of the proteins or RNAs do they interact. The research involves the analysis and aligning of multiple sequences for template selection, and the prediction of molecular models of the proteins and RNAs involved, besides the docking attempts of two models to find the possible interaction. Seven of the EBV factors were successfully generated and docked with RPL 22. Root mean square distance (RMSD) calculated between the docked proteins show strong interactions between RPL 22 and seven of the factors, which interaction may lead to arrested biological functions such as apoptosis, cell signaling and cell proliferation.

Keywords: Epstein Barr virus, nasopharyngeal cancer, ribosomal proteins L22, bioinformatics tools.

ABSTRAK

Virus Epstein Barr telah dikenal pasti daripada sel Burkitt-Limfoma. Virus ini kemudiannya ditemui dalam sel kanser nasofarinks yang membuatkan virus ini mendapat perhatian penyelidik. Gen EBV menghasilkan faktor-faktor tertentu apabila mengalami pengekspresian. Kajian terdahulu menunjukkan bahawa EBER 1, salah satu faktor virus EBV telah berinteraksi dengan protein ribosom L22 dan dipercayai telah menyebabkan kanser. Justeru, kajian ini dijalankan untuk mengenal pasti kemungkinan wujudnya perhubungan antara protein ribosom L22 dan faktor-faktor EBV yang lain menggunakan kaedah bioinformatik. Kajian ini melibatkan analisis penjajaran jujukan berganda, ramalan struktur tertier protein dan RNA serta pencantuman dua model protein untuk meramal penginteraksian. Tujuh daripada faktor-faktor EBV telah berjaya dihasilkan dan masing-masing dicantumkan dengan protein ribosom L22. Jarak RMSD telah diukur dan pencantuman protein-protein tersebut menunjukkan perhubungan kuat melalui motif protein ribosom L22 yang mungkin menyebabkan proses biologi tertentu terbantut.

Kata kunci: virus Epstein Barr, kanser nasofarinks, protein ribosom L22, alat-alat bioinformatik.

1.0 Introduction

Ribosomal protein L22 (RPL22) was known to associate with Epstein Barr virus-encoded small RNAs 1 (EBER 1) as studied by Toczyski (1993), though the molecular mechanism of the interaction has yet to be elucidated. Typically, EBV genes will produce factors such as Epstein Barr nuclear antigens (EBNA), latent membrane protein (LMP) an EBV-encoded small RNAs (EBER). However, when EBV genome are purified and expressed, tegument proteins will also be produced, which were usually found to occupy the space between the nucleocapsid and the envelope of the virus particle and may involve in transcriptional regulation, apoptosis and DNA replication (Kelly, Fraefel, Kunningham & Diefenbach, 2009). These types of proteins typically exist in all types of herpes virions, including human herpesvirus Type 4, which is also known as Epstein Barr virus. When the genome is expressed, tegument proteins such as BPLF, BOLF, BVRF, BRRF, BHLF, BMRF, BLRF, BZLF and BARF will be produced, and each of them plays their respective significant roles in functioning of herpesvirus (Johannsen *et al.*, 2004). Besides the typical factors of EBV such as EBNA, LMP and EBER, it is possible that these tegument proteins may also interact with ribosomal protein L22 (RPL22). In predicting protein-protein interaction or protein-RNA interaction, comparative modeling of the factors should be done in order to obtain the template protein model, before proceeding to next sequential step which is molecular docking. In building good protein models, Basic Local Alignment Search Tool (BLAST) will be used to find the template sequences of the proteins. In this case, position specific iterative BLAST (PSI-BLAST) which is one of the protein BLAST algorithm was used for its higher sensitivity in searching template sequences compared to standard protein BLAST. BLAST search identified seven factors of the EBV genes which templates are exist in the Protein Data Bank. Multiple sequence alignment was done to identify protein motifs and conserved regions. In building of protein model, SWISS

MODEL web server was used, and the resulting models were analyzed for QMEAN4 score and Z-score. The evaluation and validation of built protein models was done by using *PROCHECK*, a protein quality check tool that evaluate angles of protein residues and visualized them into Ramachandran plot. For RNA, energy dot plot was evaluated for its quality. After models evaluation, docking was done between RPL 22 and the seven factors of EBV, namely BARF 1, BGLF 5, BHRF 1, BZLF 2, glycoprotein B, glycoprotein H and glycoprotein L. Docking attempt was also done between RPL 22 and EBER 2.

2.0 Literature Review

2.1 Epstein Barr Virus

Epstein Barr virus is a type of herpes virus, which infection could remain latent for years without showing any symptoms or infection. This virus is predicted to be responsible for approximately 1% of human cancers, besides the possibility to contribute to other disorders such as multiple sclerosis (Arvey *et al.*, 2012). EBV was also perpetually detected in nasopharyngeal carcinoma (NPC) which tumour results from proliferation of a single cell infected with EBV (Raab-Traub, 2002). Its mechanism of infection in NPC was reported to be via SC protein interaction (Lin *et al.*, 1997). When the EBV genes are expressed, commonly produced factors are EBNA, LMP and EBER. However, according to Johannsen *et al.* (2004), when complete purified genome of EBV are translated, it will also produce tegument proteins; a type of protein that is found between nucleolus and envelope of herpes virion particles. These proteins may be associated with apoptosis, DNA replication or transcriptional regulations, besides their main function as capsid or scaffold proteins (Kelly *et al.*, 2009). The expression of EBV genes may result in production of tegument proteins such as BDLF, BPLF, BFRF, BORF, BVRF, BSRF, BBRF, BNRF, BLRF, BZLF and BALF. According to Gent *et al.* (2014), one of the proteins, namely BPLF 1 may have interfered in innate immune evasion via toll-like receptor signaling. In another study by Zhang *et al.* (1996), it was stated that BZLF1 and BMLF1 that play important roles in EBV lytic replication may interact with each other and modulate both viral transcription and translation. Thus, it is also possible that these proteins might be associated with ribosomal protein L22 (RPL22), a ribosomal protein that was earlier found to interact with EBER 1 of EBV genes by Toczyski, Matera, Ward and Steitz (1993).

2.2 Ribosomal protein L22

Ribosomal proteins has been known to involve in protein biosynthesis and are generally thought to be essential components in ribosome biogenesis, and they may be critical for both regulatory and structural functions of the ribosome (O'Leary *et al.*, 2013). Besides their function in synthesizing proteins, ribosomal proteins are believed to possess extra ribosomal functions (Wool, 1996). There has been increasing numbers of diseases that developed from mutations of genes encoding factors involved in ribosome biogenesis (Liu & Ellis, 2006). According to Toczyski and Steitz (1993), ribosomal protein L22 (RPL22) is believed to be associated with Epstein Barr virus encoded RNA 1 (EBER 1) and somehow relocalizes the protein from nucleolus into the nucleoplasm. This causes an enhanced growth potential in human Burkitt's Lymphoma cells, thus causing a tumorigenicity (Houmani, Davis & Ruf, 2009).

2.3 Bioinformatics tools

Comparative modeling is a technique of building molecular models by using a template or homologue of a particular molecule. In predicting protein-protein interaction or protein-RNA interaction, a good model is needed so that the next sequential step which is molecular docking will be a success. For a good model, it must satisfy a few criteria before it can be selected for docking (Baxevanis & Francis-Oulette, 2005). Before proceeding to comparative modeling, a target template sequence must be retrieved from the database. In this case, Uniprot KB database (http://www.uniprot.org/) for protein database and RNACentral (http://rnacentral.org/) for RNA database were used. Target sequences are retrieved together with the database ID or accession number.

2.3.1 Basic Local Alignment Search Tool (BLAST)

For templates retrieval, *Basic Local Alignment Search Tool (BLAST)* was used both for protein and nucleotide sequences. As for protein *BLAST* algorithm, position-specific iteration *BLAST (PSI-BLAST)* was used since it is more sensitive to weak but biologically relevant similarities compared to standard protein *BLAST* (Altschul *et al.*, 1997). PSI-BLAST performs iterative searches with a protein query, in which sequences found in one round of search are used to build a custom score model for the next round. *BLOSUM* series; namely *BLOSUM-62* amino acid substitution matrix was used in selection of template sequences. Altschul *et al.* (1997) also stated that *PSI-BLAST* is more refined in terms of its algorithm, where two hit method was created. This method involves two steps; the first one is the basic scanning of database for words that lies within the threshold value (*T*) when aligned with some word within the query sequence, while the second step is to check whether each hit lies within an alignment with sufficient score matrix. According to Altschul *et al.* (1997), the sensitivity of this method is evaluated by using the High Scoring Segment Pair (HSP) score. For parameters in selection of template sequences, expect value (E-value) is evaluated, which define the number of alignments expected by chance with a particular score or better. This is a default sorting metric and for *PSI-BLAST*, default E-value of 0.005 was selected, in which the template sequences with E-values higher than this threshold will not be recommended.

2.3.2 *CLUSTAL OMEGA*

CLUSTAL OMEGA (Thompson, Gibson, Plewniak, Jeanmougin & Higgins, 1997) are used for multiple alignments of proteins or nucleic acids sequences for identification of common motifs that possess certain biological functions. These tools can also be used to identify motifs in a newly characterized sequences for providing insights towards possible biological functions (Misener & Krawetz, 2000). *CLUSTAL OMEGA* is a new version of

multiple alignment tool, in which it allows alignments of proteins, RNAs and DNAs with higher accuracy and improved scaling to many sequences (McWilliam et al., 2013). As mentioned by Larkin *et al.* (2007), *CLUSTAL* programs utilize the Neighbor-Joining (NJ) method and Unweighted Pair-Group Method (UPGMA) method. The UPGMA method applies the "-clustering=UPGMA" algorithm in sequence alignments. In refining alignment, Iteration method which optimizes Weighted Sum of Pairs (WSP) score is a very effective method (Larkin *et al.*, 2007). The multiple alignments done will be viewed and edited by using *Jalview*.

2.3.3 *SWISS-MODEL*

SWISS-MODEL (http://swissmodel.expasy.org/) was initiated in 1993 by Manuel Peitsch, and is currently developed by Computational Structural Biology Group based in Swiss Institute of Bioinformatics (SIB) and Biozentrum of the University of Basel, Switzerland (Kiefer, Arnold, Kunzli, Bordoli & Scwede, 2009). This bioinformatics tool is used for predicting 3D molecular models of protein by using amino acids sequences and homology modelling techniques (Biasini *et al.*, 2014). It also provides model quality estimates by using QMEAN4 potential (Benkert, Kunzli & Schwede, 2009) and further model evaluation by using Continuous Automated Model Evaluation (CAMEO) system (http://cameo3d.org/).

2.3.4 *Mfold*

Mfold (http://mfold.rna.albany.edu/?q=mfold) was developed in late 1980s by Michael Zuker for prediction of RNA secondary structure, which delegates the algorithms of minimum free energy (ΔG) and minimum free energy for foldings that must possess any particular base pair (Zuker, 2003). Free energy minimization method is the basic principle in predicting RNA secondary structure, where the structure with minimal energy will be selected as the most acceptable (Burkowski, 2009). The pattern and distribution of free

energy will be visualized in a triangular plot called energy dot plot. Burkowski (2009) also stated that stacking energies value of RNA loops, pins and stems will be calculated for accurate prediction of RNA models. The most recent *Mfold* program is using the improved algorithms initiated by Mathews, Sabina, Zuker and Turner (1999), which applied expanded sequence dependence of thermodynamic parameters.

2.3.5 *PROCHECK*

Evaluation and quality estimation need to be done for prediction of the most acceptable models by using *PROCHECK*. This tool is hosted by European Bioinformatics Institutes (EBI) and is used to check for stereochemical quality of the modelled protein structures (Laskowski, MacArthur, Moss & Thornton, 1993). The global parameters in checking for the quality of a protein structure is the distribution of phi, psi, chi 1 torsion angles visualized in Ramachandran plots and hydrogen bond energies in the modelled structure (Morris, MacArthur, Hutchinson & Thornton, 1992). This tool also checks for 'ideal' bond length or angles for a protein model. According to Baxevanis *et al.* (2005), a good protein model must have minimum exposed hydrophobic residues and maximum polar charged residues.

2.3.6 *Cluspro 2.0*

Cluspro 2.0 (http://cluspro.bu.edu/) will be the tool used for protein-protein docking. Docking attempts of two protein molecules will be done to predict whether the involved molecules interact, and at which region or binding sites do they interact. This tool implement an algorithm which evaluate billions of putative complexes and perform selection based on complementarities (Comeau, Gatchell, Vajda & Camacho, 2004). According to Comeau *et al.* (2004), further clustering will be done after selection of structures with good electrostatic and desolvation free energy.

3.0 Research Methodology

Table 1. Bioinformatics tools and origin that were used.

Tool	Host server/Inventor
Protein Data Bank (PDB) (http://www.rcsb.org/pdb/)	Research Collaboratory for Structural Bioinformatics (RCSB)/ (Hamilton, 1971)
Uniprot KB (http://www.uniprot.org/)	European Bioinformatics Institute, (EMBL-EBI)/ (Apweiler, 2002)
GenBank (http://www.ncbi.nlm.nih.gov/genbank)	National Centre for Biotechnology Information (NCBI) (2009)
RNACentral (http://rnacentral.org/)	European Bioinformatics Institute (EMBL-EBI)/ (Bateman *et al.*, 2011)
Basic Local Alignment Search Tool (BLAST) (http://blast.ncbi.nlm.nih.gov/Blast)	National Centre for Biotechnology Information (NCBI)/ (Altschul, Gish, Miller, Myers & Lipman, 1990)
CLUSTAL OMEGA (http://www.clustal.org/clustal2/)	Swiss Institute of Bioinformatics Biozentrum, University of Basel/ (Larkin *et al.*, 2007)
SWISS-MODEL (http://swissmodel.expasy.org/)	Swiss Institute of Bioinformatics Biozentrum, University of Basel/ (Arnold, Bordoli, Kopp & Schwede, 2006)
PROCHECK (http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/)	European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI)/ (Laskowski, MacArthur & Thornton, 1993)
Swiss PDB Viewer (http://spdbv.vital-it.ch/)	Swiss Institute of Bioinformatics Biozentrum, University of Basel/ (Guex, 1994)
Mfold (http://mfold.rna.albany.edu/?q=mfold)	The RNA Institute, State University of New York/ (Zuker & Markham, 1995)
ClusPro 2.0 (http://cluspro.bu.edu/)	ClusPro server/ (Kozakov *et al.*, 2013)

Jalview (http://www.jalview.org/)	The Barton Group/ (Waterhouse, Procter, Martin, Clamp & Barton, 2009)
RasMol (http://www.RasMol.org/)	Glaxo Wellcome Research & Development/ (Bernstein, 2000)
Prosite (http://prosite.expasy.org/)	Swiss Institute of Bioinformatics Biozentrum, University of Basel (2009)
Vector Alignment Search Tool (VAST) (http://structure.ncbi.nlm.nih.gov/Structure/VAST/vast.shtml)	National Centre for Biotechnology Information (NCBI) (2014)

3.1 Data mining and collection

Data mining was done in order to obtain the sequences of proteins and nucleic acid from the databases. For protein sequences such as ribosomal protein L22 (RPL22), Epstein Barr nuclear antigen (EBNA), latent membrane protein (LMP), and other tegument proteins of Epstein Barr virus, non redundant Uniprot KB database and Protein Data Bank were used as sources. As for RNA sequences such as EBV-encoded RNA (EBER), RNACentral and GenBank were used. In retrieving the proteins, accession number or database ID of each proteins and RNAs were recorded. In order to identify the site of interaction, the protein motif of RPL22 was searched against the *Prosite* database.

3.2 Basic Local Alignment Search Tool (BLAST)

BLAST was performed to search for sequences having most significant similarities with the target sequences against Protein Data Bank and RNA databases to search for possible templates. PSI-BLAST was used for searching of protein templates by using the default parameters of E-value threshold less than 0.005 and *BLOSUM-62* substitution scoring matrix, template sequences with identity and query cover ranging from 70% to 100% and E-value closest to 0.0 were selected. E-value defines the number of distinct alignments,

with a score equivalent to or better than S (max score), that are expected to occur in database search by chance. The lower the E-value shows higher significance of the score. Max score means the highest alignment score (bit-score) between the query sequence and the database sequence segment. Moreover, query coverage defines percent of the query length that is included in the aligned segment, which is calculated over all segment. Identity, on the other hand, means the overall similarity between the query sequence with the template results. The resulted templates with query cover and identity less than 70% were considered as not having suitable templates. The selected *BLAST* templates were used to perform multiple sequence alignment using *Clustal* programs to study the alignments and motifs of protein.

3.3 Multiple sequence alignment

Multiple sequence alignment was done by using *CLUSTAL OMEGA* web server. The alignment was done in order to visualize the motif and patterns of the target proteins with their templates, as well as to compare and observe the target-template alignment for conserved regions. The multiple sequence alignment was also carried out as a prerequisite step for the next method, which is protein modeling that will be done by using *SWISS MODEL*.

3.4 Protein and RNA modeling

In protein modeling, the prediction of 3-dimensional structure of the proteins involved in the experiment was done by using the alignment mode, which requires the input of target-template alignment done earlier by using *CLUSTAL OMEGA*. The SWISS MODEL server used the models obtained from the Protein Data Bank as templates to generate protein structures that are somehow similar with the target sequence. The model scores and

properties to be analyzed such as solvation torsion angles, QMEAN4 score and Global Model Quality Estimates (GMQE) score were also provided by the SWISS MODEL web server. In RNA modeling, since the generation of RNA tertiary structure was not possible due to unstable properties of RNAs and unavailability of templates in the Nucleic Acid Database (NDB), only secondary structures were able to be generated by using *Mfold*. *Mfold* requires direct input of RNA sequences, thus target-template alignments were not needed. The resulting models were saved in pdb format and will later be validated.

3.5 Protein models validation

Validation of protein models was done to ensure that the quality of the models generated are of high quality and reliable. The quality check for protein models generated was performed by *PROCHECK* which shows the quality of the models by visualizing the protein residues into Ramachandran Plot, stereochemistry of the models, main chain and side chain parameters, bond length and angles, chi plots as well as residue properties.

3.6 Vector Alignment Search Tool (*VAST*)

VAST search was performed look for possible structural neighbours for RPL 22, and sequentially to provide a hint or to biologically prove the existence of interactions which exist between RPL 22 and the factors of EBV genes, if any. *VAST* search produces 3D structures that are somehow similar with 3D model of RPL 22. The obtained structural neighbours were then filtered for human proteins and searched against *Prosite* database of protein motifs to check for interactions, which was cross-linked to *InTact* database of protein interactions. *InTact* database then revealed the interactions which existed between the structural neighbours and other proteins, and the candidate partners were selected for EBV proteins or RNAs. If the structural neighbours do interact with EBV factors, hence it

is assumed that RPL 22 will also interact with the factors. The selected candidate partners were then proceeded to docking step.

3.7 Protein docking

The docking between candidate partners retrieved from *InTact* and RPL 22 was carried out first. After the 3D structures of proteins were validated and accepted as models of good quality, the attempts to dock the receptor protein RPL22 with respective ligands; the factors of EBV genes, were carried out . The docking was done by using *Cluspro 2.0*, a molecular docking software hosted by Swiss Bioinformatics Institute. After the docking was done, analyzation of the docked models must be carried out to calculate the probability of interaction between the receptor and the ligands. *Deepview* or *Swiss PDB Viewer* was used to visualize the docked models and calculate the root mean square distance (RMSD) between receptor's and ligands' protein residues. The interaction was observed whether they interact at a specific motif on the receptor (RPL22).

4.0 Results & Discussion

4.1 Data mining

The accession numbers of EBV proteins and RNAs were collected (refer to Appendix A) from Uniprot KB database (http://www.uniprot.org/). Protein motifs of RPL 22 were identified using *Prosite* database (http://prosite.expasy.org/).

Table 2. The possible motifs on the receptor protein RPL22. (Abbreviation: G-Glycine, K-Lysine, I-Isoleucine, D-Aspartic acid, A-Alanine, N-Asparagine, L-Leucine, V-Valine, T-Threonine, S-Serine, R-Arginine, E-Glutamic acid, Y-Tyrosine)

Motif and Function	**Protein sequence (motif)**	**Sequence numbers**
Amidation site	gGKK	11-14
N-myristoylation site	GImdAA	32-37
	GNlgGG	54-59
	GGvvTI	58-63
Protein kinase C phosphorylation site	SkR (phosphoserine)	79-81
	TkK (phosphothreonine)	87-89
Tyrosine kinase phosphorylation site	KesyElr.Y	107-114

According to *Prosite* database of protein motifs, RPL22 motifs which were recorded and presented in table 2 shows the motifs which function as the site of attachment for factors which are responsible in cell signal transduction pathway. For example, protein kinase C aids in phosphorylation of a particular amino acid for it to subsequently activate other proteins that are involved in a signaling cascade. The phosphorylation takes place on the stated motifs on RPL22. Thus, RPL22 somehow plays a significant role in regulation of cell signaling. Serine phosphorylation plays a significant role in a wide range of cellular processes. In a report by Sluss, Armata, Gallant and Jones (2004), serine phosphorylation